Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of a periplasmic glucose-binding protein from *Thermotoga maritima*

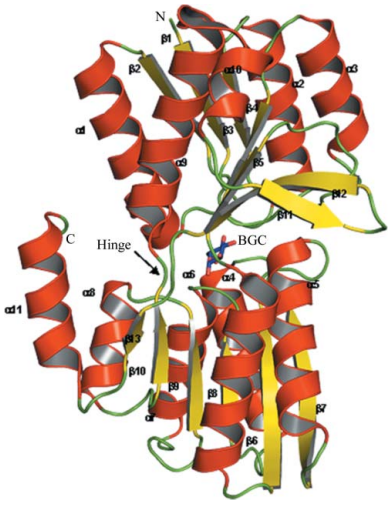
Kandavelu Palani,[a] Desigan Kumaran,[a] Stephen K. Burley[b] and Subramanyam Swaminathan[a]*

[a]Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA, and [b]Eli Lilly and Co., San Diego, CA 92121, USA

Correspondence e-mail: swami@bnl.gov

Received 22 June 2012
Accepted 1 November 2012

PDB Reference: glucose-binding protein, 2qvc



ABC transport systems have been characterized in organisms ranging from bacteria to humans. In most bacterial systems, the periplasmic component is the primary determinant of specificity of the transport complex as a whole. Here, the X-ray crystal structure of a periplasmic glucose-binding protein (GBP) from *Thermotoga maritima* determined at 2.4 Å resolution is reported. The molecule consists of two similar α/β domains connected by a three-stranded hinge region. In the current structure, a ligand (β-D-glucose) is buried between the two domains, which have adopted a closed conformation. Details of the substrate-binding sites revealed features that determine substrate specificity. *In toto*, ten residues from both domains form eight hydrogen bonds to the bound sugar and four aromatic residues (two from each domain) stabilize the substrate through stacking interactions.

1. Introduction

Transport of molecules across membrane lipid bilayers is central and critical to cell physiology, including uptake of nutrients, elimination of waste products, energy generation and cell signaling. ABC transporter proteins for the import or export of essential nutrients and other substances are widely distributed in nature (Higgins, 1992; Dean, Hamon *et al.*, 2001; Dean, Rzhetsky *et al.*, 2001) and form approximately 20–30% of the genes in a whole genome. The range of substrates that they transport varies from ions and small organic molecules to lipids, carbohydrates, amino acids, peptides and even whole proteins. Clinically relevant ABC exporters have been implicated in various pathologic conditions such as multi-drug resistance of cancer cells (Gottesman & Ambudkar, 2001; Gadsby *et al.*, 2006), cystic fibrosis and Stargardt disease (Dean, Hamon *et al.*, 2001; Dean, Rzhetsky *et al.*, 2001). Glucose-binding proteins (GBPs) belong to a large family of soluble periplasmic proteins that support distinct biological functions in Gram-negative bacteria (Mowbray *et al.*, 1990). In bacteria, this periplasmic component is the primary determinant of the specificity of the transport complex as a whole. Progression from the resting state to the transition state is driven by the engagement of substrate-loaded periplasmic binding protein rather than the binding of ATP to the Walker A and B motifs in ABC transporter proteins. Here, we report the X-ray crystal structure of a periplasmic glucose-binding protein from *Thermotoga maritima* bound to glucose at 2.4 Å resolution. *T. maritima* GBP consists of two similar structural domains connected by a three-stranded hinge region and is believed to exist in a dynamic equilibrium between closed (ligand-bound) and open (apo) conformations in solution. In our crystal structure, β-D-glucose is bound at the domain interface, where it makes extensive hydrogen-bonding and hydrophobic interactions with both domains.

2. Materials and methods

2.1. Protein production

The target gene (tm0114) for NYSGXRC target 11013q was cloned using polymerase chain reaction from *T. maritima* genomic DNA using a forward (ACCATAGGTGTTATCGGAAA) and a reverse (CTTTTATTGGAATTCCGAGTTCTT) primer. The predicted

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost shell.

Space group	*R*3
Unit-cell parameters (Å)	$a = b = 123.3$, $c = 280.6$
Method	SAD
Wavelength (Å)	0.9792
Resolution (Å)	30.8–2.40 (2.49–2.40)
No. of unique reflections	60160
Average multiplicity	9.7 (3.4)
$R_{\rm merge}$† (%)	9.6 (38.7)
Overall completeness (%)	96.0 (72.0)
$\langle I/\sigma(I)\rangle$	13.1 (2.5)
Refinement statistics	
No. of unique reflections	57916
R factor‡	0.212
$R_{\rm free}$§	0.243
No. of protein atoms	9120
No. of heteroatoms	48
No. of water molecules	287
R.m.s.d. bond lengths (Å)	0.008
R.m.s.d. bond angles (°)	1.4
Average *B* factors (Å^2)	
Protein	31.97
Ligand (BGC)	20.90
Waters	31.05
Ramachandran plot statistics, residues in (%)	
Core region	89.8
Additionally allowed region	9.4
Generously allowed region	0.8

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of symmetry-related reflections $I_i(hkl)$. ‡ *R* factor $= \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and the calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ was calculated using 5% of the data withheld from refinement.

signal peptide (1–31) was not included in cloning. The amplified gene was gel-purified and cloned into BC-pSGX3 vector designed to express the protein of interest with a C-terminal hexahistidine affinity tag. Protein expression/purification utilized previously published protocols described in detail in the PSI Knowledgebase (http://sbkb.org/tt/search?targetid=NYSGXRC-11013q&lab=NYSGXRC).

2.2. Crystallization and data collection

Initial crystallization screening of GBP was carried out with the high-throughput Index and Crystal Screens (Hampton Research) using a TECAN crystallization robot. The optimized sitting-drop vapor-diffusion crystallization involved mixing 2 µl protein solution (21 mg ml^{-1}) with 2 µl reservoir solution consisting of 0.2 *M* ammonium sulfate, 30% PEG 4000 and equilibrating against 650 µl of the same reservoir solution. Good-quality crystals were flash-cooled in liquid nitrogen after transferring them into mother liquor containing 15–20%(*v*/*v*) glycerol. X-ray diffraction data were collected under standard cryogenic conditions on beamline X29A at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. Data were integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The cystals belonged to space group *R*3, with unit-cell parameters $a = b = 123.3$, $c = 280.6$ Å, $\gamma = 120.0^\circ$ (hexagonal setting). The calculated Matthews coefficient was 2.98 Å^3 Da^{-1}, assuming the presence of four molecules per asymmetric unit, which corresponds to an estimated solvent content of ∼60%. Data-collection and refinement statistics are provided in Table 1.

2.3. Crystal structure determination

The crystal structure was determined at 2.4 Å resolution *via* the single-wavelength anomalous dispersion (SAD) technique using selenomethionine-substituted protein. The selenium subsubstructure was determined using *SHELXD* (Sheldrick, 2008) and revealed all 11 possible Se positions per molecule. *SHARP* (de La Fortelle & Bricogne, 1997) was used for phase refinement, and density modification was performed with *SOLOMON* (Abrahams & Leslie, 1996). The resulting phases were used for automated model building with *ARP*/*wARP* (Perrakis *et al.*, 1999), which yielded ∼90% of the polypeptide chain. The model-building tool *O* (Jones *et al.*, 1991) was used to build the remaining polypeptide chain. All four molecules (*A*, *B*, *C* and *D*) in the asymmetric unit were then generated using noncrystallographic symmetry (NCS) operators derived from the Se positions. Well defined residual electron density was observed in all four molecules and could be modeled as β-D-glucose even though no ligand was added during crystallization. The atomic model was refined to convergence with the *CNS* (Brünger *et al.*, 1998) slow-cool annealing method alternating with manual model building (final *R* factor and $R_{\rm free}$ values of 0.21 and 0.24, respectively). The final atomic model consisted of 1204 residues, four β-D-glucose molecules and 287 water molecules. Model validation with *PROCHECK* (Laskowski *et al.*, 1993) revealed that 89.8% of the residues were in the most favorable region and 9.4% were in the generously allowed region of the Ramachandran diagram. All figures were drawn using *PyMOL* (DeLano, 2002). Refined atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org; PDB entry 2qvc). While the refinement of this crystal structure was in progress, the crystal structure of an identical protein with bound β-D-glucose was published in a different space group (Tian *et al.*, 2007). These two structures agree with an r.m.s.d. of 0.34 Å for 297 equivalent pairs of C$^\alpha$ atoms.

3. Results and discussion

3.1. The overall structure

The asymmetric unit of our *T. maritima* GBP crystals contained four structurally similar protomers (pairwise C$^\alpha$ root-mean-square deviations of 0.15–0.26 Å), which occur as monomers in solution as judged from size-exclusion chromatography (data not shown). Each protomer has approximate dimensions of 55 × 35 × 35 Å and is composed of two similar Rossmann-fold domains (pairwise r.m.s.d. of ∼1.6 Å for 74 aligned C$^\alpha$-atom pairs; Fig. 1). Residues 32–132 and 267–308 form domain 1 and residues 133–266 and 309–333 form domain 2. Each domain consists of an α/β-fold with a core of β-sheet flanked by layers of α-helices on both sides (Fig. 1). Domain 1 consists of a seven-stranded β-sheet (β1–5 and β11–12) with three α-helices (α1 and α9–10) on one side of the sheet and two α-helices (α2–3) on the other side. Domain 2 includes six β-strands (β6–10 and β13) with two α-helices (α4–5) lying on one side of the sheet and four α-helices (α6–8 and α11) on the opposite face.

The domains are connected by a hinge made up of three segments (132–134, 265–267 and 308–312) of the polypeptide chain, with the ligand-binding site formed by the cleft between the domains. Sugar binding proceeds from an encounter complex with the open form of the protein, which is thought to predominate in the absence of ligand (Zou *et al.*, 1993), followed by hinge bending to give the closed form observed in our cocrystal structure.

3.2. The binding cleft

The sugar-binding site is buried within the interface between the two domains. β-D-Glucose (chair conformation) makes extensive hydrogen-bonding interactions with both domains. The location of this sugar-binding site is similar among structurally characterized periplasmic sugar-binding proteins (Zou *et al.*, 1993; Borrok *et al.*, 2007). The side chains of residues Lys39, Trp45, Asn167, Arg171,

Table 2
Residues forming the binding cleft in β-D-glucose-binding protein (2qvc) and the corresponding residues in *E. coli* RBP (2dri) and *S. typhimurium* GBP (3gbp).

2qvc	2dri	3gbp
Hydrogen-bond donor/acceptor residues		
Arg171 (O1, O2)	Asp89 (O1, O2)	Asp154 (O1)
Gln266 (O2)	Arg141 (O2, O3)	Asn91 (O5, O6)
Asp246 (O1)	Gln235 (O2)	Arg158 (O2)
Asn167 (O3)	Asp215 (O3, O4)	Asn256 (O1)
Ala221 (O5)	Asn190 (O4)	Asp236 (O2, O3)
Lys39 (O3)	Asn13 (O4)	Asn211 (O3)
Trp45 (O4)		Asp14 (O4)
Hydrophobic residues forming the binding surface		
Tyr44	Phe15	Phe16
Trp45	Phe16	His152
Tyr220	Phe164	Trp183
Tyr222		

Asp246 and Gln266 and the main-chain N atom of Ala221 form an intricate network of hydrogen bonds with the sugar-ring hydroxyl groups and the sugar-ring O atom (Fig. 2*a*). In the binding cleft, most of the polar groups are hydrogen-bonded to the sugar molecule. In addition to the interacting polar residues, four aromatic groups (Tyr44 and Trp45 from domain 1 and Tyr220 and Tyr222 from domain 2) make hydrophobic contacts with the bound ligand (Fig. 2*b*). β-D-Glucose is stacked between residues Tyr44 and Tyr220, while Trp45 lies perpendicular to the Tyr44 ring. The precise configuration of the aromatic groups in the binding cleft is thought to be an important determinant of binding specificity (Chaudhuri *et al.*, 1999).

3.3. The hinge region

Three short polypeptide-chain segments (linking the domains) together provide a functionally important hinge that support inter-domain conformational changes. The major differences between the unliganded and glucose-bound structures of GPBs lie within the three hinge segments (Fig. 1). The first hinge segment (residues Asp308 and Asp312) connects β12 from domain 1 and β13 from domain 2. The second hinge segment (Gly265 and Arg267) links β9 from domain 2 with α9 from domain 1. Gly132 and Asp134 of the third hinge segment connect β5 from domain 1 and α4 from domain 2. Movement in the hinge region must be coordinated among these three segments and the relative domain motion is vital for correct recognition of the membrane component for transport (Björkman & Mowbray, 1998). *HINGEprot* (Emekli *et al.*, 2008) analysis identified three hinge residues (Asp134, Gln266 and Gly310).

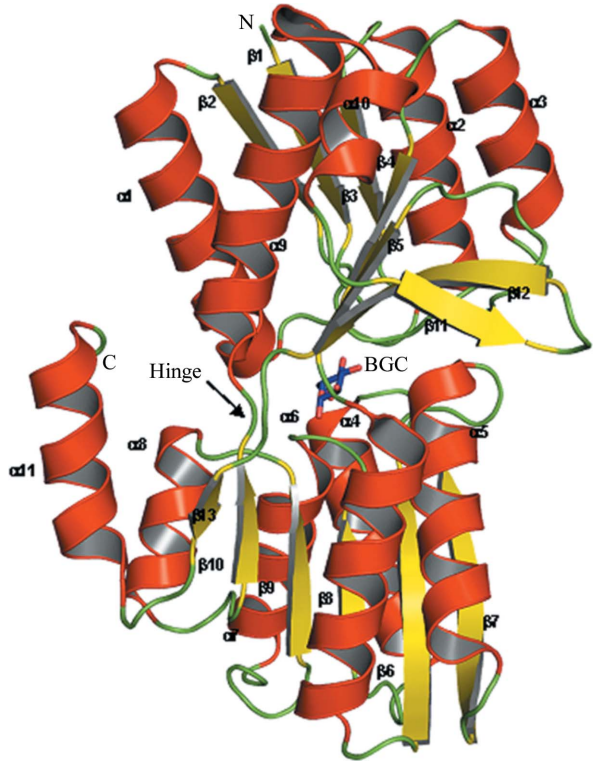


Figure 1
Cartoon representation of the overall structure of GBP, with the bound β-D-glucose shown as a stick figure. The hinge region is indicated by an arrow and secondary-structural elements are labeled.

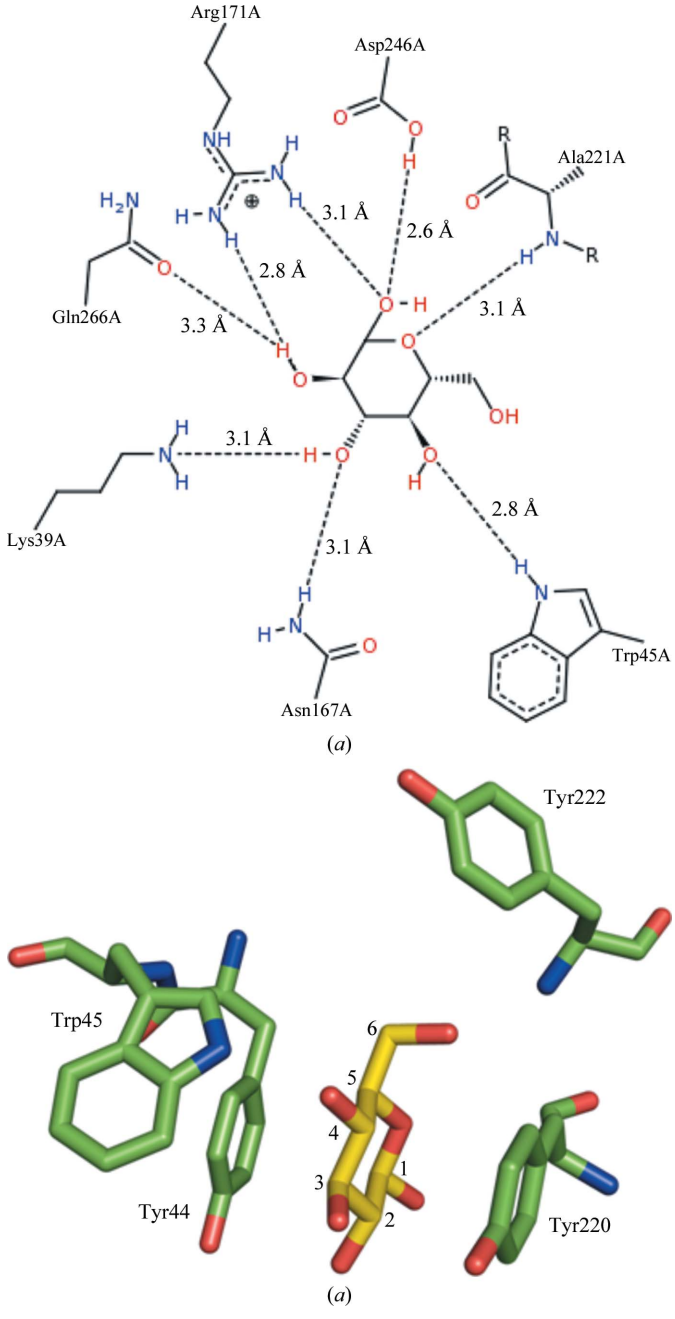


Figure 2
(*a*) Hydrogen-bonding interactions between sugar and GBP. H atoms were positioned stereochemically. However, the hydrogen-bond distances are between the heavy atoms. (*b*) Sugar–aromatic interactions.

```
2QVC   -MSLTIGVIGKSVHPYWSQVEQGVKAAGKALGVDTKFFVPQKEDINAQLQMLESFIAEGV 59
3GBP   ADTRIGVTIYKYDDNFMSVVRKAIEKDGKSAPDVQLLMNDSQNDQSKQNDQIDVLLAKGV 60
2DRI   -KDTIALVVSTLNNPFFVSLKDGAQKEADKLG-YNLVVLDSQNNPAKELANVQDLTVRGT 58

2QVC   NGIAIAPSDPTAVIPTIKKALEMGIPVVTLDTDSPDSGRYV-----YIGTDNYQAGYTAG 114
3GBP   KALAINLVDPAAAGTVIEKARGQNVPVVFFNKEPSRKALDSYDKAYYVGTDSKESGVIQG 120
2DRI   KILLINPTDSDAVGNAVKMANQANIPVITLDRQATKGEVVS-----HIASDNVLGGKIAG 113

2QVC   LIMKELLGGKG----------KVVIGTGSLTAMNSLQFIQGFKDAIKDSEIE--IVDILN 162
3GBP   DLIAKHWQANQGWDLNKDGKIQYVLLKGEPGHPDAEARTTYVVKELNDKGIQTEQLALDT 180
2DRI   DYIAKKAGEGA----------KVIELQGIAGTSAARERGEGFQQAVAAHKFN--VLASQP 161

2QVC   DEEDGARAVSLAEAALN--AHPDLDAFFGVYAYNGPAQALVVKNAGKVGKVKIVCFDTTP 220
3GBP   AMWDTAQAKDKMDAWLSGPNANKIEVVIANNDAMAMGAVEALKAHNKS-SIPVFGVDALP 239
2DRI   ADFDRIKGLNVMQNLLT--AHPDVQAVFAQNDEMALGALRALQTAGKS-DVMVVGFDGTP 218

2QVC   DILQYVKEGVIQATMGQRPYMMGYLSVTVLYLMNKIGVQNTLMMLPKVKVDGKVDYVIDT 280
3GBP   EALALVKSGAMAGTVLNDANNQA---KATFDLAKNLAEGKGAADGTSWKIENKIVRVPYV 296
2DRI   DGEKAVNDGKLAATIAQLPDQIG---AKGVETADKVLKGE--KVQAKYPVDLKLVVKQ-- 271

2QVC   GVDVVTPENLDEYLKKMEELGIPIKEG 307
3GBP   GVDKDNLSEFT---------------- 307
2DRI   ------------------------- 
```

Figure 3
Multiple sequence alignment of similar structures. Residues involved in hydrogen-bond interactions are marked in red in the present structure and in light blue in the comparative structures.

3.4. Comparison of the binding sites with similar structures

The residues that are found in the binding cleft of GBP from *T. maritima* (PDB entry 2qvc) are similar to those found in GBP from *Salmonella typhimurium* (PDB entry 3gbp; Mowbray *et al.*, 1990) and in *E. coli* ribose-binding protein (RBP; PDB entry 2dri; Björkman *et al.*, 1994). Of the seven residues that make contact with β-D-glucose in 2qvc, six equivalent residues interact with the ribose in 2dri, while seven equivalent residues are observed to bind glucose in 3gbp (Fig. 3). In 2qvc, the number of hydrophobic residues forming the binding surface exceeds those observed in the other two proteins (Table 2). Although none of these hydrophobic interactions are conserved, the general characteristics of the binding site are very similar. A superposition of these three cocrystal structures is depicted in Fig. 4.

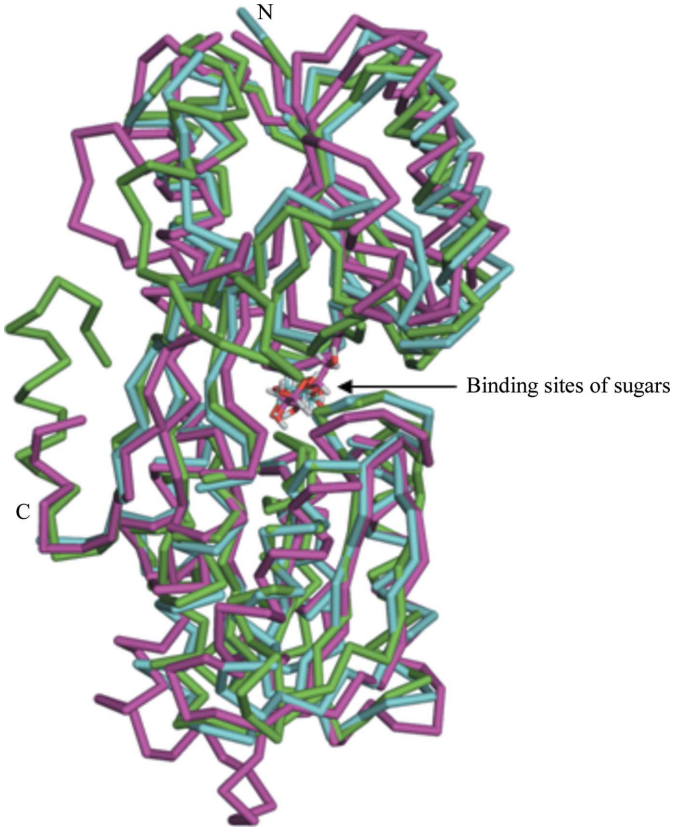


Figure 4
C^{α} superposition of 2qvc (green), 2dri (cyan) and 3gbp (magenta) showing a common closed sugar-bound conformation.

4. Conclusions

The overall crystal structure of GBP bound to β-D-glucose is comparable with those of other periplasmic sugar-binding proteins from Gram-negative bacteria, all of which exhibit a closed ligand-bound conformation. The characteristic features of this protein subfamily, such as domain structure and organization, the tripartite hinge region and the location of the binding cleft within the domain interface, are all similar. Both domains participate in ligand binding in the closed liganded structures, burying the bound ligand. β-D-Glucose binds to *T. maritima* GBP in its lowest energy chair form, which has also been observed in other members of this subfamily. Hydrogen-bonding and hydrophobic interactions of the sugar with both domains of the protein stabilize the closed form. In summary, these interactions both substitute for water binding to this highly soluble class of ligands and exploit aromatic sugar stacking, which is thought to contribute to binding.

This research was supported by a U54 award from the National Institute of General Medical Sciences to the NYSGXRC (GM074945; PI SKB) under DOE Prime Contract No. DEAC02-98CH10886 with Brookhaven National Laboratory. We gratefully acknowledge data-collection support from beamline X29 at the National Synchrotron Light Source.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Björkman, A. J., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A. & Mowbray, S. L. (1994). *J. Biol. Chem.* **269**, 30206–30211.
Björkman, A. J. & Mowbray, S. L. (1998). *J. Mol. Biol.* **279**, 651–664.

Borrok, M. J., Kiessling, L. L. & Forest, K. T. (2007). *Protein Sci.* **16**, 1032–1041.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chaudhuri, B. N., Ko, J., Park, C., Jones, T. A. & Mowbray, S. L. (1999). *J. Mol. Biol.* **286**, 1519–1531.
Dean, M., Hamon, Y. & Chimini, G. (2001). *J. Lipid Res.* **42**, 1007–1017.
Dean, M., Rzhetsky, A. & Allikmets, R. (2001). *Genome Res.* **11**, 1156–1166.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R. & Haliloglu, T. (2008). *Proteins*, **70**, 1219–1227.
Gadsby, D. C., Vergani, P. & Csanády, L. (2006). *Nature (London)*, **440**, 477–483.
Gottesman, M. M. & Ambudkar, S. V. (2001). *J. Bioenerg. Biomembr.* **33**, 453–458.
Higgins, C. F. (1992). *Annu. Rev. Cell Biol.* **8**, 67–113.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–493.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mowbray, S. L., Smith, R. D. & Cole, L. B. (1990). *Receptor*, **1**, 41–53.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Tian, Y., Cuneo, M. J., Changela, A., Höcker, B., Beese, L. S. & Hellinga, H. W. (2007). *Protein Sci.* **16**, 2240–2250.
Zou, J., Flocco, M. M. & Mowbray, S. L. (1993). *J. Mol. Biol.* **233**, 739–752.